



TITLE:

待ち行列における最適状態の提案 (待行列理論とその応用)

AUTHOR(S):

吉岡, 良雄; 中村, 維男; 重井, 芳治

CITATION:

吉岡, 良雄 ...[et al]. 待ち行列における最適状態の提案 (待行列理論とその応用). 数理解析研究所講究録 1981, 425: 146-161

ISSUE DATE:

1981-04

URL:

<http://hdl.handle.net/2433/102602>

RIGHT:

待ち行列における最適状態の提案

岩手大・工 吉岡良雄

東北大・工 中村維男, 重井芳治

1. 序 論

電話網における交換局, 計算機のジョブ処理および計算機網などは, 待ち行列問題として解析されている^{[1]-[3]}。しかし, その評価測度については, 種々異なる。窓口(処理系)の利用者にとって, 待ち時間が非常に長いことは, 非常に損失である。また, 窓口の利用率を高くすることも必要である。このことから, 待ち行列において, 利用者の待ち時間と窓口の利用率を共に考慮した評価量の提案が必要であると思われる。このような評価量に関しては, 1977年, 著者ら^[4]によって, 単一処理系待ち行列における実効処理量を提案している。また, 1978年, A. Gieschlar^[5] & L. Kleinrock^[6]によって, 実効処理量と同じような評価量であるパワー関数を, 計算機網の解析などに導入している。

本論文では, 文献[4]に示した待ち行列の最適状態を,

複数窓口に拡張した待ち行列の最適状態（最適トラヒック）を提示する。これは、利用者の通過時間をできるだけ小さくすること、かつ処理系の利用率をできるだけ高くすることから求められる。また、ふくそうの一般的定義を示し、この最適状態と待ち行列系内のふくそうとの関連について述べる。そして、これらの例として、 $M/E_k/1$ 待ち行列、 $M/G/1$ 待ち行列、 $M/M/S$ 待ち行列における最適状態を示す。

2. 待ち行列の最適条件

一般的な場合を考え、図1に示すように、利用者の到着は平均到着率 λ の一般分布、利用者の処理時間分布は平均処理率 μ の一般分布、同じ処理系が S 個、行列が無限に可能なモデルである $G/G/S$ 待ち行列を考える。そして、一利用者のもつ平均処理量（例えば、情報の大きさなど）を ρ とし、利用者の系内における平均待ち時間を w_q とする。

一方、計算機網のように、網内を通過する時間が評価の対象となる場合を想定し、利用者の待ち時間も処理時間と考えれば、図2に示すような行列のない S 個の仮想的な処理系のモデルとなる。そして、この仮想的な処理系のモデルが、図1に示す待ち行列モデルと等価であるとすれば、一仮想的な処理系の平均処理率 μ' は、次式で与えられる。

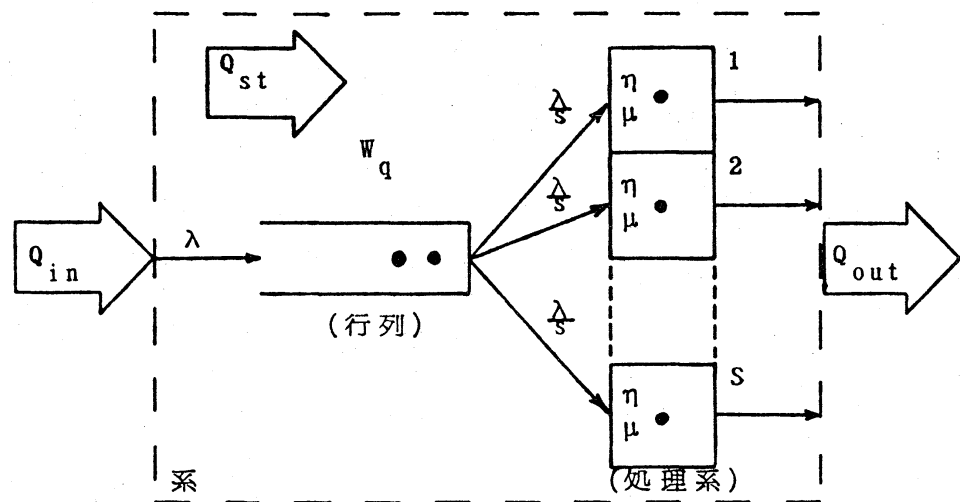


図1 複数の処理系をもつ待ち行列モデル

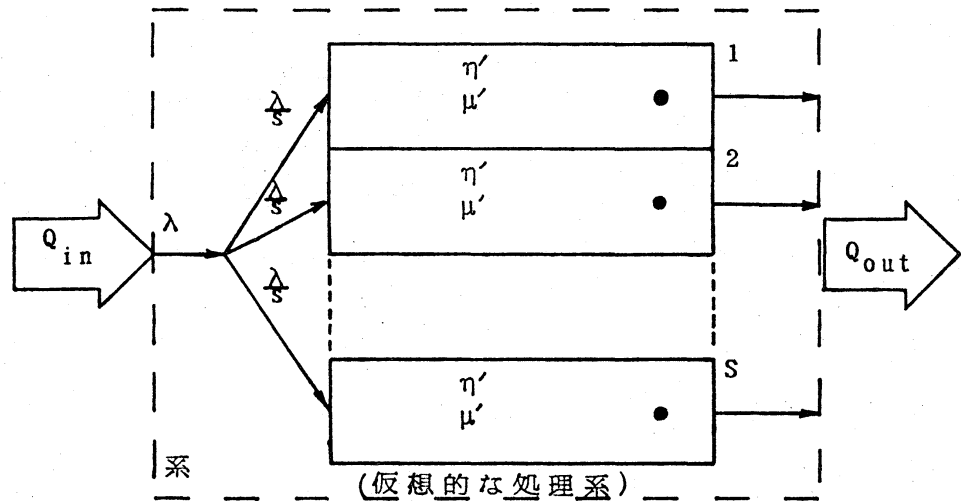


図2 仮想的な処理系のモデル

$$\mu' = \frac{1}{w_q + \frac{1}{\mu}} \quad (1)$$

図1および図2の各モデルをそれぞれ一つの系と考え、利用者のもつ処理量の流れとして、この系に流入する処理量を Q_{in} 、この系から流出する処理量を Q_{out} 、および系内に存在

する処理量を Q_{st} とする。また、定常状態の系を考えれば、各処理量に対し次式が成立する。

$$\beta = \frac{d}{dt} Q_{in} = -\frac{d}{dt} Q_{out} = \text{一定} \quad (2)$$

$$\frac{d}{dt} Q_{st} = 0 \quad \text{または} \quad Q_{st} = L \cdot g \quad (3)$$

ここで、 β は、系に単位時間に流入する処理量であり、流入率とよぶ。また、この量で系から単位時間に流出することから、これを系の処理率ともよぶ。 L は、系の長さである。

さらに、十分長い時間 $T(Q)$ の間に処理量 Q が系に流入し、処理されたとすれば、定常状態を考えているから、図1における処理系の利用率 γ 、および図2における仮想的な処理系の利用率に相当する量 γ' は、それぞれ次式となる。

$$\gamma = \frac{Q}{S \cdot g} \cdot \frac{1}{\mu} / T(Q) = \frac{\lambda}{S \mu} = \rho \quad (4)$$

$$\gamma' = \frac{Q}{S \cdot g} \cdot (W_g + \frac{1}{\mu}) / T(Q) \quad (5)$$

ここで、 ρ は、トラヒック密度である。

これより、次の関係式を得る。

$$\gamma' = (\mu \cdot W_g + 1) \cdot \gamma \quad (6)$$

$$\begin{aligned}\beta &= \lambda \cdot \bar{g} = \frac{Q}{T(Q)} = S \cdot \eta' \cdot v^* \\ &= S \cdot \eta \cdot v^* \cdot \mu \cdot W_g + S \cdot \eta \cdot v^* \quad (7)\end{aligned}$$

ここで、 v^* は、利用者の系内における平均速度（実効速度とよぶ）であり、次式で定義される。

$$v^* = \frac{\bar{g}}{W_g + \frac{1}{\mu}} \quad (8)$$

式 (7) において、 $S \cdot \eta \cdot v^* \cdot \mu \cdot W_g$ は、利用者の待ち時間による部分であり、 $S \cdot \eta \cdot v^*$ は、利用者の実効的な処理の部分である。そこで、これらをそれぞれ、無効処理率、有効処理率（文献 [4] の実効処理率と同じ）とよぶことにする。

以上のことから、待ち行列の評価量は、 $\eta \cdot v^*$ で与えられ、これを最大にする待ち行列の状態は、最適となる。

3. 電気回路への対応による待ち行列の最適条件の導出

図1において、利用者が系に到着し出るまでの時間を T_g とすれば、これは次式で与えられる。

$$T_g = W_g + \frac{1}{\mu} \quad (9)$$

単位時間に処理系を利用する利用者数は、 λ/S であるか

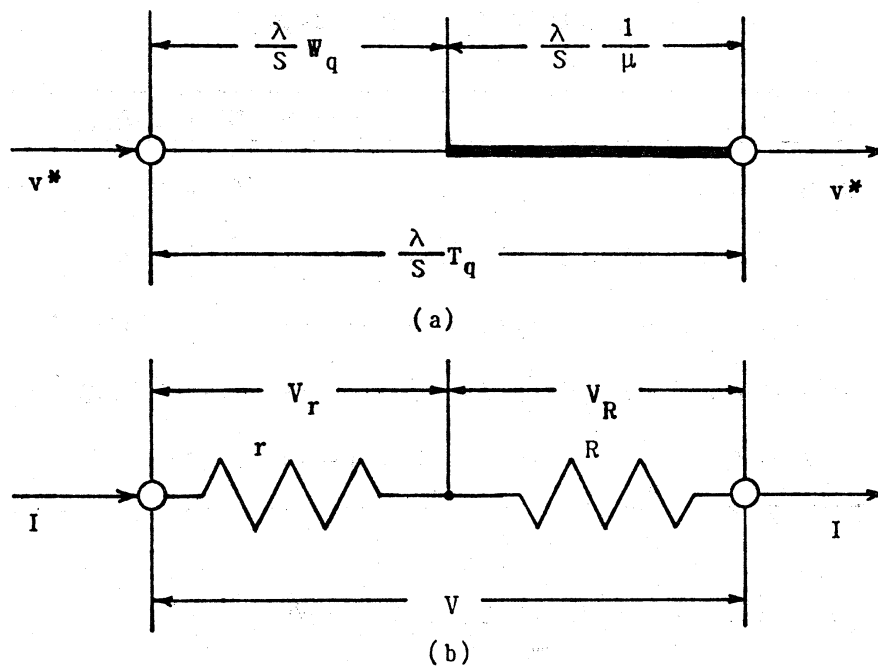


図3 待ち行列 (a) と電気回路 (b) との対応

ら、この利用者数について式 (9) を考えれば、次式を得る。

$$\frac{\lambda}{S} T_q = \frac{\lambda}{S} W_q + \frac{\lambda}{S\mu} \quad (10)$$

一方、系内を流れる利用者のもつ処理量の平均速度 v^* は、式 (8) で与えられる。これらを図に示せば、図3 (a) のようになる。また、図3 (b) は、非線形な抵抗 r , R を直列に接続し、その両端に V の電圧を加えた電気回路である。これから明らかなように、表1に示す対応関係が成立する。

この対応関係から、電気回路の性質を用い、非線形な抵抗 r , R に消費する電力に相当する量 P_r , P_R は、それぞれ次式となる。

表 1 待ち行列と電気回路との対応関係

待ち行列		↔	電気回路	
処理量	q	↔	電荷量	q
利用者数 λ/S の所用時間	$\frac{\lambda}{S} T_q$	↔	端子電圧	V
利用者数 λ/S の損失時間	$\frac{\lambda}{S} W_q$	↔	r の両端電圧	V_r
利用者数 λ/S の処理時間	$\frac{\lambda}{S} \frac{1}{\mu}$	↔	R の両端電圧	V_R
実効速度	v^*	↔	回路電流	I

$$P_r = I \cdot V_r = v^* \cdot \frac{\lambda}{S} W_q = \gamma \cdot v^* \cdot \mu \cdot W_q \quad (11)$$

$$P_R = I \cdot V_R = v^* \cdot \frac{\lambda}{S} \frac{1}{\mu} = \gamma \cdot v^* \quad (12)$$

以上から、 $\gamma \cdot v^*$ は、電気回路を電流と電圧を共に考慮した評価量である電力に相当し、待ち行列の評価量とすることが出来る。よって、これを最大にすることは、物理的に意味がある。

4. 待ち行列内のふくそうについて

電話交換などにおいては、呼びの接続サービスが不可能となる状態をふくそうとよび、これを確率で示しふくそう率とよんでいる。また、待ち行列では、到着した呼（利用者）が処理されない状態（ $\rho \geq 1$ ）でふくそうが発生しているとさ

れている。しかし、ふくそうの一般的定義は、系内の利用者の混雑状態を意味するものである。このことから、トラヒック密度 ρ が $1 > \rho > 0$ であっても、ふくそうは発生していることになる。そこで、定常状態の系における定量的なふくそうの発生の定義をここで示す。

式(5)(6)の ρ' は、仮想的なものであり、 $\rho' > 1$ となり得る。 $\rho' > 1$ であることは、仮想的な処理系で利用者のすべてを処理できないことを意味している。これから、 ρ' は、系内の混雑状態を示すパラメータと考えられ、待ち行列における処理系の負荷は、 $\rho' > 1$ で重く、 $1 > \rho' \geq 0$ で軽いといえる。そこで、特に $\rho' > 1$ となる系の状態をふくそうが発生していると定義し、 ρ' における系の状態は式(7)より次の関係となる。

$$\begin{aligned} \beta/s &> v^* & (\rho' > 1), \\ \beta/s &= v^* & (\rho' = 1), \\ \beta/s &< v^* & (1 > \rho' \geq 0) \end{aligned} \quad (13)$$

ここで、 β/s は、一処理系において単位時間に処理する平均の処理量である。また、 v^* は、利用者の一処理系における実効速度である。よって、ふくそうの発生しない上限($\rho' = 1$)では、 $\beta/s = v^*$ となる。

一方, $S=1$ のとき, 式 (6) から, ρ は系の平均長に等しい。このことから, ふくそうの発生しない上限では, 系の平均長は1である。また, このとき, 系の処理率 ρ と系の実効速度 \bar{v}^* は, 等しくなる。

4. 待ち行列の最適状態の例

単一処理系待ち行列の最適状態の例として, $M/E_k/1$ 待ち行列および $M/G/1$ 待ち行列の場合を取り扱う。また, 複数処理系待ち行列の最適状態の例として, $M/M/S$ 待ち行列の場合を取り扱う。そして, それぞれの場合について, 数値計算を行い, ふくそうとの関連を示す。

(1) $M/E_k/1$ 待ち行列の最適状態

この場合における平均待ち時間 W_q は, 次式で与えられる。

$$W_q = \frac{(k+1) \cdot \rho^2}{2 \cdot k \cdot \lambda \cdot (1-\rho)} \quad (14)$$

よって, 有効処理率 $\rho \cdot \bar{v}^*$ は, 次式となる。

$$\rho \cdot \bar{v}^* = \frac{2 \cdot \mu \cdot k \cdot (1-\rho) \cdot \rho}{2 \cdot k - (k-1) \cdot \rho} \quad (15)$$

これを図に示せば, 図4のようになる。図から $\rho \cdot \bar{v}^*$ は, トラヒック密度 ρ に対し最大値をとる。このときのトラヒック密度 ρ の値を求めるため, 式 (15) を ρ で微分し零とおくこ

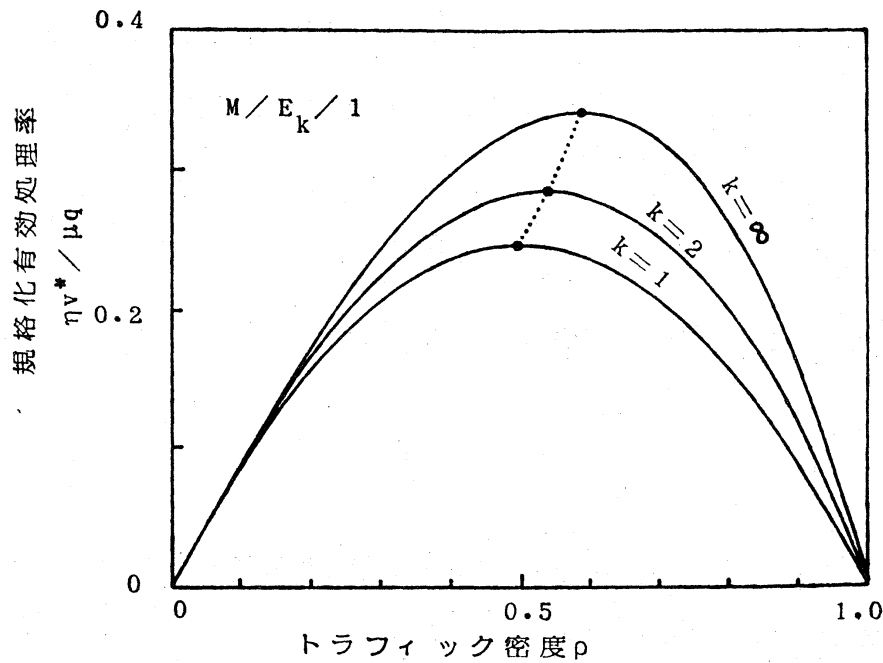


図4 トラフィック密度 ρ に対する規格化有効処理率 $\eta \cdot v^* / \mu \cdot q$ の変化

とによって、次式を得る。

$$(k-1) \cdot \rho^2 - 4 \cdot k \cdot \rho + 2 \cdot k = 0 \quad (16)$$

この解で、 $1 > \rho > 0$ かつ $\eta \cdot v^*$ を最大にするトラヒック密度を ρ_m とおけば、次式となる。

$$\rho_m = \begin{cases} \frac{1}{2} & (k=1) \\ \frac{2 \cdot k - \sqrt{2k(k+1)}}{k-1} & (k \neq 1) \\ 2 - \sqrt{2} & (k=\infty) \end{cases} \quad (17)$$

このとき, $\rho' = 1$ となり, 定義から系内にふくそうが発生していない。特に $k=1$ のとき, 平均待ち時間 W_g と平均処理時間 $1/\mu$ が等しくなるという興味ある結果となる。

(2) M/G/1 待ち行列の最適状態

この場合, 平均待ち時間 W_g は, 次式で与えられる^[8]。

$$W_g = \frac{\lambda}{2 \cdot (1-\rho)} \cdot \int_0^{\infty} x^2 \cdot D(x) dx \quad (18)$$

ここで, $D(x)$ は, 処理分布の確率密度関数である。

よって, 有効処理率 $\rho \cdot \eta^*$ は, 次式となる。

$$\rho \cdot \eta^* = \frac{2 \cdot \mu \cdot g \cdot \rho \cdot (1-\rho)}{2 \cdot (1-\rho) + \rho \cdot \mu^2 \cdot \int_0^{\infty} x^2 \cdot D(x) dx} \quad (19)$$

これを ρ で微分すれば, 次式を得る。

$$\left\{ 1 - \frac{\mu^2}{2} \cdot \int_0^{\infty} x^2 \cdot D(x) dx \right\} \rho^2 - 2 \cdot \rho + 1 = 0 \quad (20)$$

この解で $1 > \rho > 0$ かつ $\rho \cdot \eta^*$ を最大にするトラヒック密度 ρ_m は, 次式となる。

$$\rho_m = \frac{1}{1 + \sqrt{\frac{\mu^2}{2} \cdot \int_0^{\infty} x^2 \cdot D(x) \cdot dx}} \quad (21)$$

このとき, $\rho' = 1$ となり, 定義から系内にふくそうが発生

していない)。

(3) M/M/S待ち行列の最適状態

この場合, 平均待ち時間 W_q は, 次式で与えられる^[7]。

$$W_q = \frac{\rho}{\lambda \cdot (1-\rho)} \cdot \pi \quad (22)$$

ここで, π は, 処理系がすべて塞がる確率を示し, 次式で与えられる。

$$\pi = \frac{S^S \cdot \rho^S}{S! \cdot (1-\rho)} \left/ \left\{ \sum_{n=0}^{S-1} \frac{S^n \rho^n}{n!} + \frac{S^S \cdot \rho^S}{S! \cdot (1-\rho)} \right\} \right. \quad (23)$$

よ, ρ , 有効処理率 $\eta \cdot \mu^*$ は, 次式となる。

$$\eta \cdot \mu^* = \frac{S \cdot \mu \cdot \rho \cdot (1-\rho)}{\pi + S \cdot (1-\rho)} \quad (24)$$

また, η' は, 次式となる。

$$\eta' = \frac{\rho}{S \cdot (1-\rho)} \cdot \pi + \rho \quad (25)$$

式 (24), (25) を図に示せば, 図5, 6のようになる。図から明らかなように, 有効処理率 $\eta \cdot \mu^*$ は, トラヒック密度 ρ に対し最大値をとる。そして, η' は, $\eta \cdot \mu^*$ が最大値をとるとき, $\eta' \leq 1$ である。よ, ρ , 定義から, 系内にふくそうが発生していない。また, このとき $\mu^* \geq \beta/S$ となる。

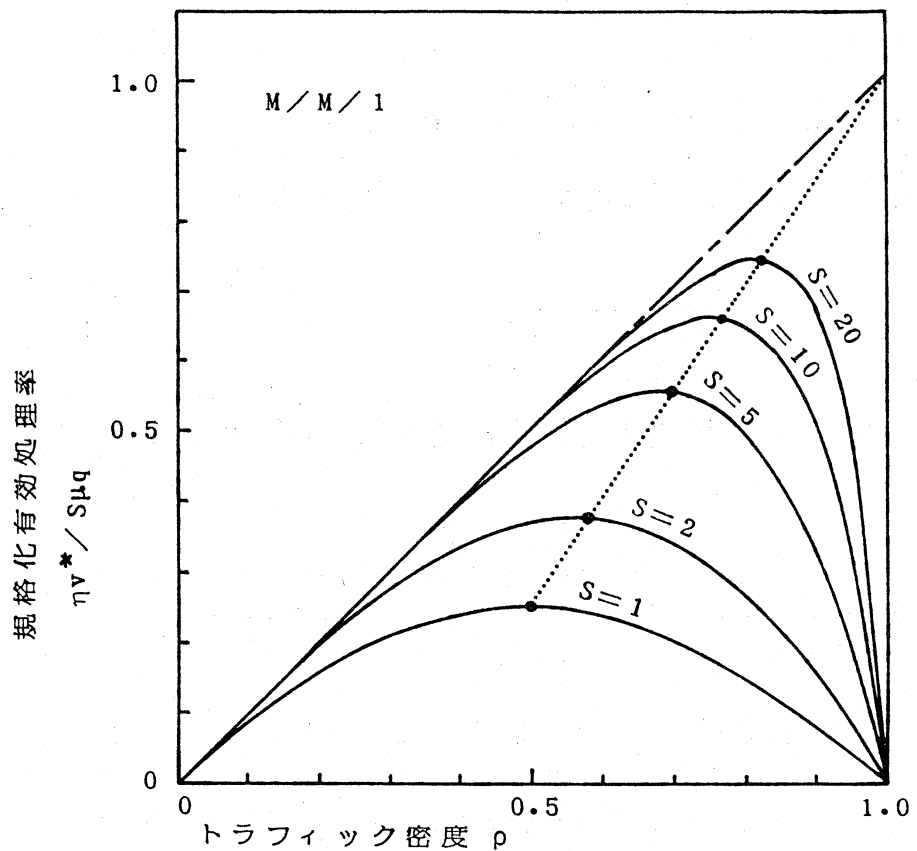


図5 トラフィック密度 ρ に対する規格化有効処理率 $\eta v^* / S \mu q$ の変化

以上、待ち行列の最適状態の例として、 $M/E_k/1$ 待ち行列、 $M/G/1$ 待ち行列、 $M/M/S$ 待ち行列の場合を取り扱った。これらから明らかなように、待ち行列の平均待ち時間が与えれば、待ち行列の最適状態が求められる。

5. 結 論

待ち行列において、利用者の待ち時間は、利用者にと、て損失である。そこで、計算機網のように、利用者の網内通過

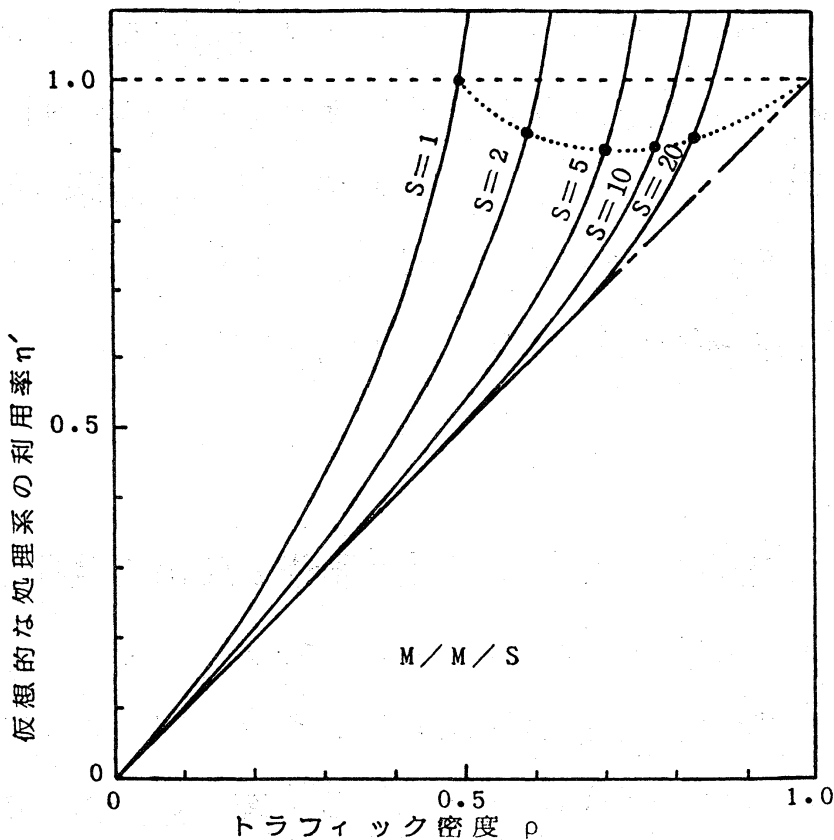


図6 トラフィック密度 ρ に対する仮想的な処理系の利用率 η' の変化

時間および網の利用率（またはスループット）が評価の対象となる場合において、この通過時間をできるだけ小さくすること、かつ処理系の利用率をできるだけ高くすることから、待ち行列の最適状態を導出した。この評価量は、利用者の実効速度と処理系の利用率の積で与えられ、これを有効処理率とよぶ。これを電気回路に対応して考えた場合、電流と電圧の積である電力に相当し、これを最大にすることは、物理的に意味がある。そして、待ち行列における有効処理率が最大

になるとき，待ち行列内のトラヒックは最適となる。また，待ち行列内のふくそうパラメータを提示し，待ち行列の最適状態との関連について示した。この待ち行列の最適状態の例として， $M/E_k/1$ 待ち行列， $M/G/1$ 待ち行列， $M/M/S$ 待ち行列の場合を取り扱い，検討を行った。

〔文 献〕

- [1] 藤木："通信トラヒック理論の基礎と応用"，信学誌，55-8，PP. 1033 - 1039，(昭47-08)。
- [2] 坂田，野口，大泉："タイムシェアリングシステムのラウンドロビン方式について"，信学論(C)，51-C，9，PP. 409 - 416，(昭43-09)。
- [3] L.G. Roberts："The ARPA computer network"，in *Computer Communication Networks*，edited by Abramson and Kuo，Prentice Hall (1972)。
- [4] 吉岡，中村，佐藤："待ち行列の最適化について"，信学論(B)，J60-B，8，PP. 590-591，(昭52-08)。
- [5] A. Giessler et al.，"Free Buffer allocation - An investigation by simulation"，*Comput. Networks*，vol. 2，PP. 191-208，1978。
- [6] L. Kleinrock："On flow control in computer network" in *Proc. Int. Conf. Commun.*，June 1978。

- [7] 本間: "待ち行列の理論", 理工学社, (1966).
- [8] 森村, 大前: "応用待ち行列理論", 日科技連, (1975)
- [9] 吉岡, 中村, 佐藤: "待ち行列の最適化とその応用",
信学技報, CST 77-16, (1977-05).